Binaural Speech Recognition in Noise and the Effect of Context

A Senior Honors Thesis

Presented in partial fulfillment of the requirements for graduation
with research distinction in Speech and Hearing Science in the undergraduate colleges
of The Ohio State University

by

Anna Fry

The Ohio State University

June 2010

Project Advisor: Dr. Christina Roup, Department of Speech & Hearing Science

Abstract

Listening and understanding speech in noisy environments is a situation that many people encounter in their daily lives. This problem is exacerbated by old age or the presence of a hearing loss. The purpose of the present study was to investigate two components that contribute to successful word recognition in noise. The first component is the advantage of binaural over monaural listening and the second is the role of contextual cues. Ten normal hearing young adults aged 20-24 years old participated. Sentences from the Speech Perception in Noise (SPIN) test were presented to the subject through insert ear phones simultaneously with multitalker babble as background noise under three conditions: 1) noise and signal presented to the right ear, 2) noise and signal presented to the left ear, and 3) noise and signal presented to both the left and right ears. The subject was told to repeat the last word in each sentence and the responses were scored as correct or incorrect. Results show that subjects perform better on word recognition tasks in the binaural listening condition versus the monaural listening conditions. Performance was also better with high predictability sentences in which the context can be useful for word recognition. Literature has shown that older adults also rely on context for word recognition, so it can be expected that for the older adults who are hard of hearing word recognition would be most successful when listening with two ears and when presented with high contextual information.

Table of Contents

Page

Acknowledgements

I would like to thank Dr. Christina Roup, Dr. Lawrence Feth, and Dr. Michelle Bourgeois for their continued support and encouragement throughout this entire project.

Chapter 1

Introduction and Literature Review

Listening and understanding speech in noisy environments is a situation that many people encounter in their daily lives. Understanding a speaker in the presence of background noise may be a difficult task for certain people as it is exacerbated by old age or the presence of a hearing loss. There are two factors that affect a person's ability to understand speech in noisy conditions: 1) binaural listening and 2) context. Binaural listening, or listening with two ears, is superior to listening with only one ear (i.e. monaural listening) for speech understanding, especially in the presence of background noise for most listeners (Brooks, 1984). This is known as the binaural advantage. Context also helps to improve speech understanding in noise. Performance is better when the listener is presented with sentences that are highly predictable and provide contextual cues versus sentences with little or no contextual information (Hutchinson, 1989). Knowledge of the topic of conversation and the presence of other contextual words provides information that helps a person fill in the gaps and determine speech where some of the acoustic information may have been lost due to background noise.

While the effects of binaural listening and context on speech understanding in noise have previously been studied as separate factors, the present study seeks to investigate the role of both. It is hypothesized that when a normal hearing young adult is listening in a noisy environment, the optimal condition for speech understanding will be when the person is able to listen with two ears and when presented with high contextual information.

Binaural Advantage

The ability to hear a speech signal with both ears instead of just one has been shown to result in better speech recognition, especially in the presence of background noise. In a study that required subjects to identify nonsense syllables that were spoken in the presence of background noise, it was found that young adults performed better when they were able to listen binaurally versus monaurally (Helfer, 1992). This data refers to a binaural advantage because of the better performance that resulted from the subject's ability to listen with two ears (binaural) as opposed to only one ear. Therefore, if a person has a hearing loss in only one ear (unilateral hearing loss), it will be more difficult for them to understand speech in a noisy environment as they do not benefit from a binaural advantage because they are only receiving substantial acoustic information in one ear. Keys (1947) also found that the best advantage for binaural listening occurs under the condition when both ears are equally stimulated. The implications of these findings are that in order to provide a person with an optimal condition for binaural listening in noise, it is important that both ears are able to receive the same amount of acoustic information. In the case of a person with any degree of hearing loss in one of their ears, it is therefore important to consider the advantage of providing some kind of acoustic amplification for the impaired ear. In the case of a person with hearing loss in both ears that are not of equal degrees, it is important to consider fitting the person with amplification on both ears instead of only on one ear. Fitting people with amplification in this manner allows them to have the best opportunity for hearing with a binaural advantage in a noisy environment.

Context

Sentences used as a part of normal conversation usually carry a great deal of contextual information which is available to the listener. Hutchinson (1989) found that both old and young adult listeners are able to recognize speech better with high predictability sentences which provide the listener with a lot of contextual information. Context, therefore, was shown to aid in a person's ability to understand speech, especially in the presence of background noise, as it helps fill in any information that was missed from the acoustic signal. A listener's speech recognition performance consists of the listener first using only sensory information to identify the individual elements of the speech stimulus and then using contextual information to substitute any elements that were missed from the acoustic information (Bronkhorst, Bosman, & Smoorenburg, 1993). A person's ability to use contextual information to aid in speech recognition, however, relies on the individual's knowledge of the language and their ability to use this knowledge to fill in the gaps from information missed from the acoustic signal (Grant & Seitz, 2000). Contextual information, therefore, is only helpful if the listener has a prior experience and understanding of the language and how it is used in order to apply previous knowledge to the new situation.

Speech Recognition and the Aging Process

Deterioration of the sense of hearing accompanies the normal aging process. Presbycusis, hearing loss caused by aging, affects the structures of the inner ear and results in a sensorineural hearing loss. Based on the structure of the ear that is affected there are several different types of presbycusis that can occur: sensory, strial, and neural. The general characteristics seen in people with presbycusis are decreased hearing sensitivity, difficulty understanding speech in noisy environments, and reduced ability to localize the source of sounds (Gates & Mills, 2005). These characteristics occur as the age-related hearing loss progresses. Usually over the age of 60 years, hearing loss begins in the high frequencies and it becomes especially difficult to understand speech in background noise (Gates & Mills, 2005). Other parts of speech become difficult to understand as the hearing loss progressively affects more areas of the hearing organ which are correlated to various other frequencies of the incoming acoustic signal. For example, when the hearing loss reaches the frequencies of the acoustic signal from 2-4 kHz, it affects the person's ability to understand voiceless consonants such as t, p, k, f, s, and ch (Gates & Mills, 2005). This deficit causes important acoustic information to be left out of the sound signal that reaches the listener. For this reason, people with presbycusis report that not only is the level of what they hear softer, but they also cannot understand what they are able to hear because many components of the signal are lost and the person must guess at the missing information in order to decipher the message (Gates & Mills, 2005).

Prevalence of age-related hearing loss

The relationship between the effects of the aging process and hearing loss are clear. Cruickshanks et al. (1998) found that 45.9% of adults ages 48-92 who participated in a study in Beaver Dam, Wisconsin had a hearing loss. In the group of people 80 years and older, the prevalence of hearing loss increased to 89.5% (Cruickshanks, et al., 1998). In the context of this study, hearing loss is defined as having a Pure Tone Average greater than 25 dB, from the 500, 1000, 2000, and 4000 Hz frequencies, in at least one ear, thus including people with both unilateral and bilateral hearing loss (Cruickshanks, et al., 1998). For the purpose of the current study it is important to include those people with both unilateral and bilateral hearing loss because any degree of hearing loss will affect the individual's ability to understand speech in noise and may keep them from benefiting from the binaural advantage.

Aging and the binaural advantage

With the normal aging process comes the deterioration of the hearing organs, causing a hearing loss to progress as a person gets older. It becomes much more difficult to hear and understand speech especially in competitive listening environments, such as background noise. However, even in these difficult listening environments, older adults exhibit a significant binaural advantage (Grose, 1996). In fact, old and young adults with normal hearing or minimal hearing loss receive the same amount of benefit from binaural listening in syllable understanding tasks, which supports the idea of a binaural advantage (Helfer, 1992). When the elderly are being fit with hearing amplification, this binaural advantage should be a factor for deciding between binaural or monaural amplification.

Aging and the effect of context

There is a great deal of evidence that shows older adults have a slower ability to process acoustic signals (Gordon-Salant, Yeni-Komshian, & Fitzgibbons, 2008). This population, however, does have the ability to make use of contextual information, especially when listening to speech in noisy conditions, despite slower processing abilities (Pichora-Fuller, Schneider, & Daneman, 1995). It has been shown that older adults can take an almost equal amount of benefit from contextual information as young adults when listening to normal conversation in the presence of background noise. Hutchinson (1989) found that 77% of older adults were able to use the contextual information as well as the young adults.

Significance

The aim of this study is to determine the most effective conditions for a normal hearing young adult to understand speech when listening in noise, both in regards to binaural listening and contextual cues. It can also be expected, however, that for an older adult or a person with a hearing loss, the ability to listen equally with both ears would increase speech understanding performance. Previous literature has shown that older adults gain an equivalent benefit from contextual cues as younger listeners in speech recognition tasks (Dubno, Ahlstrom, & Horwitz, 2000). It can therefore be expected that speech recognition would be most successful for aging and hearing impaired individuals under the same conditions as young adults with normal hearing.

In regards to context, the results of this study will also be beneficial for the speaker in a conversation in a noisy environment. If contextual cues are shown to improve a listener’s ability to understand speech, then strategies could be developed by the speaker in order to place more emphasis on the use of context in the conversation, therefore improving the amount of information that is understood in the exchange, even if some of the acoustic signal is lost in the background noise.

Chapter 2

Methods

Subjects

Ten young adults (9 female and 1 male) were recruited to participate in the present study. The subjects ranged in age from 20-24 years with a mean age of 21.7 years. All subjects had normal hearing defined as audiometric thresholds (250-8000 Hz) ≤20 dB HL and ≤10 dB air-bone gaps. Additional inclusion criteria included: 1) normal otoscopy and 2) normal tympanometry. All subjects were native speakers of American English.

Materials

Speech recognition performance was evaluated using the Speech Perception in Noise (SPIN) test (Bilger, Nuetzel, Rabinowitz & Rzeczkowski, 1984). Lists 1-8 of the SPIN sentences were used. Bilger et al. (1984) modified the original SPIN test to ensure that each list of words yielded equivalent speech recognition performance results. List equivalency allows for a different list to be used for each subject and each condition, ensuring that the results are not be affected by the difficulty of the specific set of sentences. Each list consists of a set of 50 sentences presented in background noise, specifically multitalker babble. SPIN sentences are categorized as either high predictability (HP) or low predictability (LP) based on the amount of contextual information and the predictability of the final word of the sentence. Twenty-five of the sentences in each list are categorized as HP meaning that they contain contextual information that would help the subject determine the final word in the sentence. The other 25 sentences are categorized as LP because they contain little or no contextual information. For example, a HP sentence would be “All the flowers were in *bloom*” and a LP sentence would be “I had a problem with the *bloom*” (Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984).

Procedure

Speech recognition performance in multitalker babble was measured under three response conditions: 1) monaural right, 2) monaural left, and 3) binaural. A different list of SPIN sentences was used for each response condition so that the subject heard a total of 50 sentences in each response condition (25 HP and 25 LP). The specific lists of SPIN sentences and the order of monaural and binaural response conditions were counterbalanced across subjects. The SPIN sentences were presented from a CD player (Sony CE375) through a two-channel audiometer (Grason Stadler, Model 61) at 50 dB HL via insert earphones. The multitalker babble was presented at 42 dB HL, resulting in a +8 dB signal-to-noise ratio (SN). Subjects were then instructed to repeat the final word of each sentence and the responses were scored as either correct or incorrect. Each subject participated in a single one-hour session and was paid for their time. All audiometric and experimental testing was conducted in a sound attenuating booth. Both the audiometer and tympanometer were calibrated according to the appropriate American National Standards Institute standards (ANSI, 1987, 2004).

Chapter 3

Results

Group Results

Mean SPIN recognition performance across the three response conditions is presented in Table 3.1. As expected, results revealed superior SPIN performance in the binaural condition relative to the two monaural conditions (i.e., a binaural advantage). Specifically, mean performance was 84.4% binaurally, 78.2% for the right ear, and 80.6% for the left ear. Superior binaural SPIN recognition performance can be seen for both sentence types as well as combined across sentence types, as shown in Figure 3.1. In order to determine if there were significant differences in performance due to response condition, the percentage SPIN scores were transformed to rationalized arcsine units (Studebaker, 1985) and subjected to a series of t-tests of means. There were no significant differences between the mean monaural right and monaural left SPIN scores for any of the response conditions. Therefore, for all future analyses, monaural performance was collapsed across ears (i.e., monaural combined) for comparison with binaural performance. Results revealed significant differences in SPIN performance between monaural combined and the binaural response condition. Specifically, binaural SPIN performance was significantly better than monaural combined for both the LP sentences (t_{18} = 2.6; $p < .05$) and the overall scores (t_{18} = -3.1; $p < .05$). The difference between monaural combined and binaural SPIN performance was not significant for the HP SPIN sentences.

Table 3.1: Mean recognition performance (percent correct) and standard deviations for monaural right, monaural left, and binaural response conditions across all SPIN sentence types: high predictability (HP), low predictability (LP) and overall (HP + LP).

	SPIN Performance (in percent)		
	High Predictability	Low Predictability	Overall (HP + LP)
Response Conditions			
Monaural Right			
Mean	96.0	60.4	78.2
SD	3.3	9.5	5.0
Monaural Left			
Mean	94.0	67.2	80.6
SD	6.3	10.9	6.7
Binaural			
Mean	97.2	71.6	84.4
SD	2.7	6.1	2.8

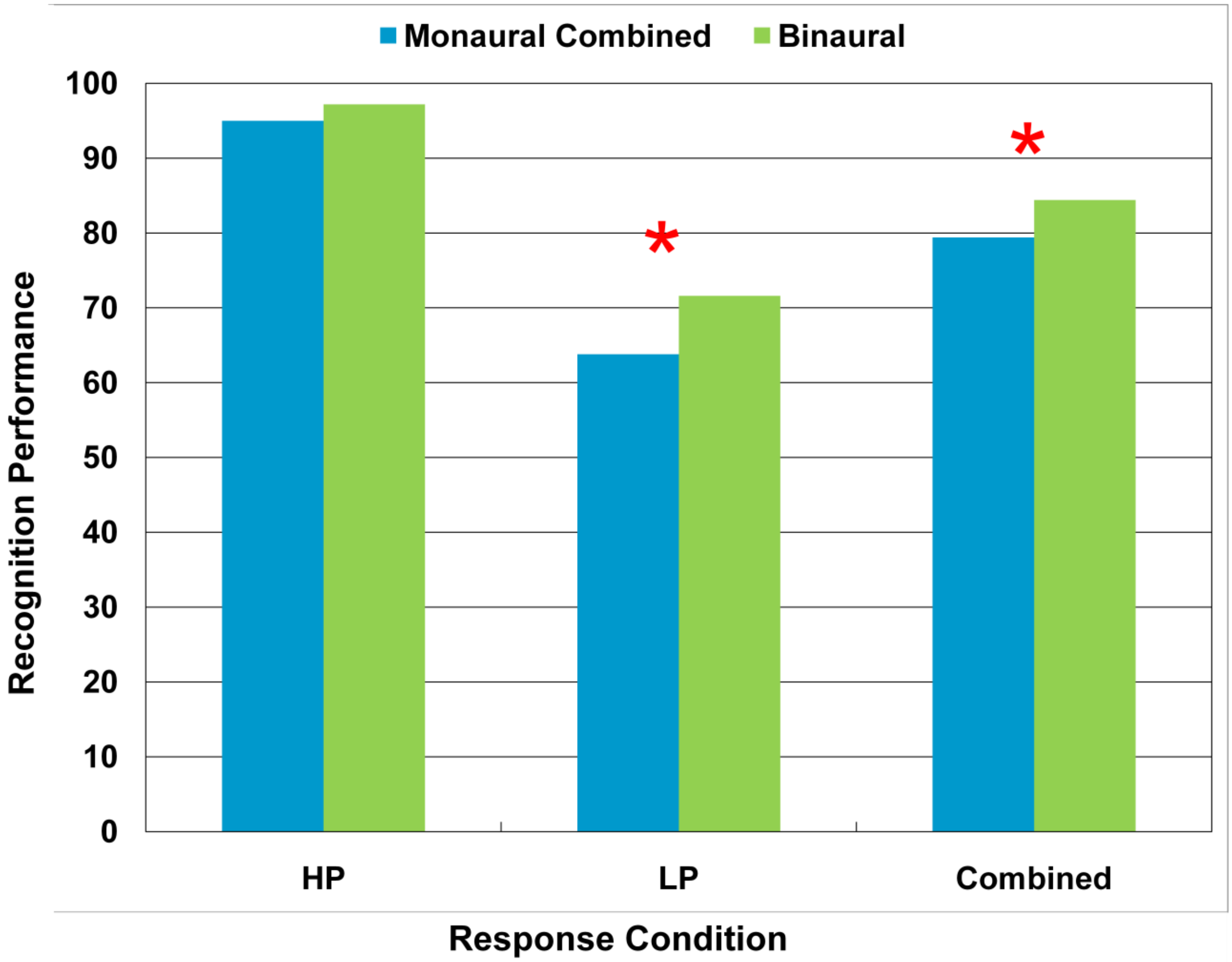

Figure 3.1: Mean recognition performance (percent correct) for monaural combined and binaural response conditions across SPIN sentence types: high predictability (HP), low predictability (LP) and combined (HP+LP).

* significant difference (p< .05)

Figure 3.2 presents mean SPIN recognition performance as a function of sentence type. Specifically, Figure 3.2 compares HP to LP SPIN sentences, demonstrating the effect of context. The effect of context on recognition performance is striking. Results revealed better SPIN performance for HP sentences than for LP sentences in both monaural combined and binaural conditions. Monaurally, average performance was 95% for HP sentences and 63.8% for LP sentences. Binaurally, average performance was 97.2% for HP sentences and 71.6% for LP sentences. In order to determine if the differences between HP and LP SPIN recognition performance were significant, the transformed data were analyzed via a t-test of means. Results revealed significant differences in recognition performance as a function sentence type. Specifically, recognition performance was significantly better for HP sentences than LP sentences for both monaural combined ($t_{18} = 12.0$; $p < .05$) and binaural ($t_{18} = 11.4$; $p < .05$) response conditions.

Individual Results

Figure 3.3 presents individual data as a bivariate plot with monaural combined recognition performance plotted on the abscissa and binaural recognition performance plotted on the ordinate. Data points that fall above the diagonal line indicate better performance in binaural condition (i.e., a binaural advantage), whereas data points that fall below the diagonal line indicate better performance in the monaural combined condition. Data points on the line indicate equal performance. As can be seen in Figure 3.3, for the individual subjects speech recognition performance in multitalker babble is best under both binaural and HP conditions. In both the HP and LP

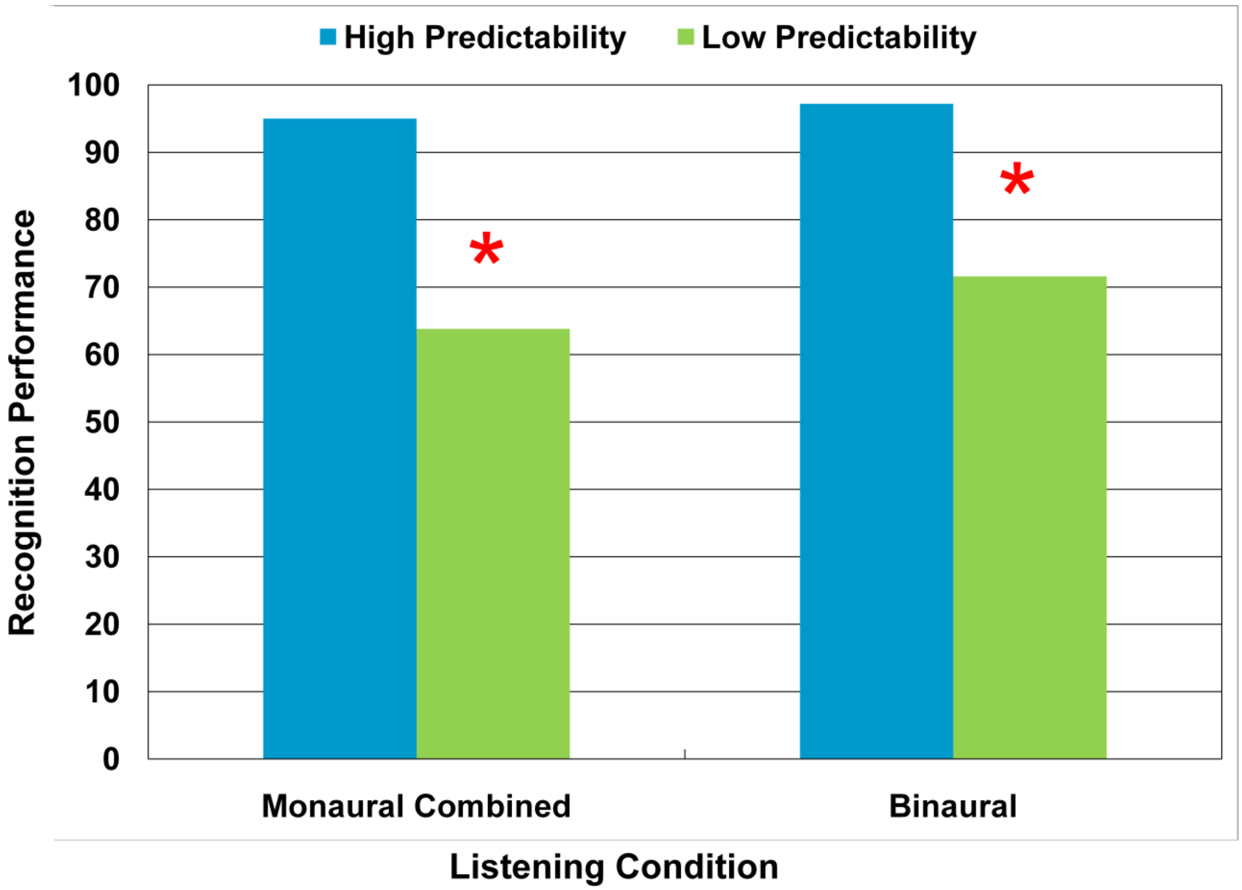

Figure 3.2: Mean recognition performance (percent correct) as a function of context: high predictability (HP) vs. low predictability (LP) SPIN sentences. Data are plotted for both monaural combined and binaural conditions.

* significant difference (p< .05)

conditions, speech recognition performance was better when listening binaurally versus monaurally (more data points above the line). In both the binaural and monaural conditions, individuals scored higher on the speech recognition task for HP sentences versus LP sentences.

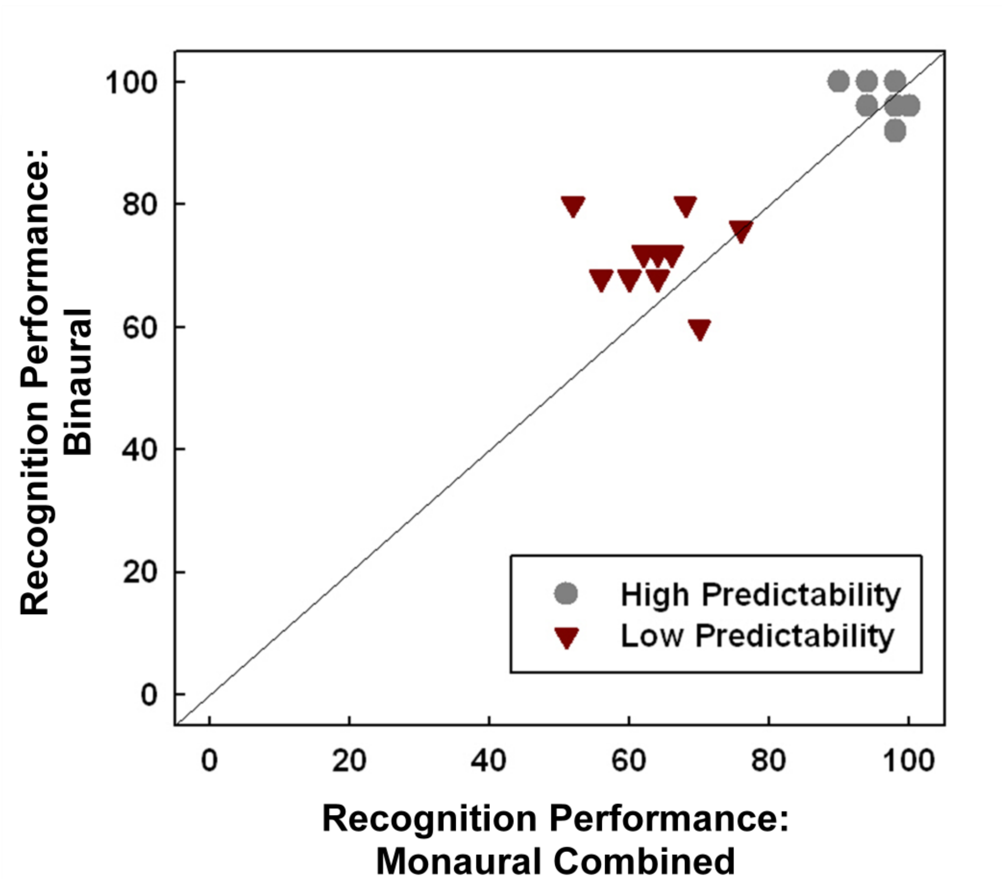

Figure 3.3: Bivariate plot of individual data with monaural combined recognition performance on the abscissa and binaural recognition performance on the ordinate. Data points above the line indicate better performance in the binaural condition (i.e., a binaural advantage), whereas data points below the line indicate better performance for monaural combined. Data points on the line indicate equal performance.

Chapter 4

Discussion

Understanding a speaker in the presence of background noise can be a difficult task due to the loss of some acoustic information in the noise. For young adults with normal hearing, the data shows that for word recognition in noisy conditions there is an advantage to listening with both ears (i.e., binaural advantage). Context also plays a role in word recognition in noise. Specifically, contextual information aids in speech understanding in difficult listening situations. Therefore, for this population, in order to understand speech in a noisy environment, the optimal listening condition is listening with two ears and being provided with as much contextual information as possible. It is interesting to note, however, that the results show that the effect of context outweighs the subjects' ability to listen with two ears. When listening to speech in noise, the difference between listening to speech with high and low contextual information resulted in a much larger difference in speech recognition performance than did the difference between binaural and monaural listening conditions.

SPIN sentences in this study were presented at a +8 dB SN. While this condition led to significant results for the LP sentences and overall SPIN sentences, performance on the HP sentences were nearly at ceiling in both the monaural and binaural conditions and there was no significant difference between these two listening conditions. While it was clear that there was a binaural advantage in the LP conditions and for the overall sentence list, no definite conclusions could be made about a binaural advantage in the HP sentences. An interesting option would be to present this same task at several different SNs. Using a SN that is closer to or below zero, meaning that the level of presentation of the multitalker babble would be equal to or more than the level of presentation of the sentences, would make word recognition a more difficult task. More conclusions may be able to be drawn from these results about a binaural advantage in situations with high contextual information.

Conclusions

Previous literature has shown that older adults exhibit a binaural advantage and also gain an equivalent benefit from contextual cues as younger listeners in speech recognition tasks (Dubno, Ahlstrom, & Horwitz, 2000). It can therefore be expected that speech recognition would be most successful for aging and hearing impaired individuals when given the ability to listen with two ears and presented with high contextual information. An area of future research for this study would be to look at the performance of older adults or people with hearing loss on the same word recognition tasks in order to compare the results to those of the young adult group.

References

American National Standards Institute. (1987). *Specifications for Instruments to Measure Aural Acoustic Impedance and Admittance (Aural Acoustic Immittance)* (ANSI S3.39-1987 R2002). New York: American National Standards Institute.

American National Standards Institute. (2004). *Specifications for Audiometers* (ANSI S3.6-2004). New York: American National Standards Institute.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research , 27*, 32-48.

Bronkhorst, A. W., Bosman, A. J., & Smoorenburg, G. F. (1993). A model for context effects in speech recognition. *J Acoust Soc Am , 93* (1), 499-509.

Bronkhorst, A. W., Brand, T., & Wagener, K. (2002). Evaluation of context effects in sentence recognition. *J Acoust Soc Am , 111* (6), 2874-2886.

Brooks, D. N. (1984). Binaural benefit- when and how much? *Scand Audiol , 13*, 237-241.

Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. K., Klein, R., Mares-Periman, J. A., et al. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin: The epidemiology of hearing loss study. *American Journal of Epidemiology , 148* (9), 879-886.

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2008). Binaural advantage for younger and older adults with normal hearing. *Journal of Speech, Language, and Hearing Research , 51*, 539-556.

Dubno, J. R., Ahlstrom, J. B., & Horwitz, A. R. (2000). Use of context by young and aged adults with normal hearing. *J Acoust Soc Am , 107* (1), 538-546.

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *www.thelancet.com , 366*, 1111-1120.

Gordon-Salant, S., Yeni-Komshian, G., & Fitzgibbons, P. (2008). The role of temporal cues in word identification by younger and older adults: Effects of sentence context. *J Acoust Soc Am , 124* (5), 3249-3260.

Grant, K. W., & Seitz, P. F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *J Acousti Soc Am , 107* (2), 1000-1011.

Grose, J. H. (1996). Binaural performance and aging. *J Am Acad Audiol , 7*, 168-174.

Haggard, M. P., & Hall, J. W. (1982). Forms of binaural summation and the implications of individual variability for binaural hearing aids. *Scand Audiol Suppl , 15*, 47-63.

Helfer, K. S. (1992). Aging and the binaural advantage in reverberation and noise. *Journal of Speech and Hearing Research , 35*, 1394-1401.

Hutchinson, K. M. (1989). Influence of sentence context on speech percpetion in young and older adults. *Journal of Gerontology , 44* (2), 36-44.

Keys, J. W. (1947). Binaural versus monaural hearing. *The Journal of the Acoustical Society of America , 19* (4), 629-631.

Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology , 47*, S72-S82.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am , 97*, 593-608.

Studebaker, G.A. (1985). A rationalized arcsine transform. *Journal of Speech and Hearing Research*, *28*, 455-462.

Wingfield, A. (1996). Cognitive factors in auditory performance: Context, speed of processing, and constraints of memory. *J Am Acad Audiol , 7*, 175-182.